ChinaXiv合作期刊

新兴高性能计算行业应用及发展战略

冯圣中[1*]　李根国[2]　栗学磊[1]　齐富民[1]　黄　典[1]　万　艺[1]　吴金成[1]

1 国家超算深圳中心　深圳　518055

2 上海超级计算中心　上海　201203

摘要　数据密集型新兴行业应用快速发展，是近年来高性能计算应用日益广泛和深入的主要特征。新兴高性能行业应用，在高性能计算系统技术创新、计算环境创新与应用创新等各个层面，都带来了新的挑战与机遇。文章在系统总结领域应用进展的基础上，概括了新兴行业应用的技术特点与挑战，提出了加大高性能计算系统核心技术创新力度、构建面向新型应用的高性能计算环境、大力推进高性能应用软件研发、大力推进传统应用的新方法开发，以及大力推进大数据人工智能等新领域基准评测工具研发等发展战略建议。

关键词　新兴行业应用，数据驱动，数据密集型计算，高性能计算

DOI　10.16418/j.issn.1000-3045.2019.06.005

chinaXiv:202303.10253v1

近 20 年来，高性能计算应用的广度前所未有地扩展，以数据驱动或数据密集型计算为主要特征的高性能计算应用不断涌现，人工智能技术在新型和传统高性能计算应用领域获得成功实践，应用领域遍及生物信息与生命科学领域、智慧城市与城市治理、网络信息安全等。这些应用又反过来对高性能计算技术，包括矩阵并行求解技术、高性能大数据处理技术、智能芯片技术等，产生巨大的影响，促进了高性能计算技术创新。

1 新兴高性能计算应用快速发展

新兴高性能计算行业应用主要为数据驱动型应用，应用领域从智慧城市到生物信息处理，乃至基本粒子物理学、天文学和宇宙学等众多领域。数据密集型计算为其最主要的计算行为特征。

1.1 高性能计算在智慧城市领域获得广泛应用

2008 年，IBM 提出智慧地球愿景，智慧城市的概念也随之出现。交通拥堵、环境污染、气候变化、

*通讯作者

资助项目：科技部重点专项（2018YFB0204400）
修改稿收到日期：2019年6月11日

chinaXiv:202303.10253v1

犯罪率上升等一系列问题降低了城市的宜居、宜产程度。智慧城市利用各类传感器采集的数据，汇聚到高性能计算平台，在集成处理的基础上，服务城市决策、社会生产和居民生活。而遍布城市的各类传感器及物联网设备采集的数据，呈现爆炸式增长。2018年，中国的数据总量约为7.6 ZB（1 ZB=10^{21}字节），占全球总量的23.4%。预计到2025年，中国的数据总量将增长至48.6 ZB，占全球数据总量的27.8%，中国将成为全球最大的数据圈[1]。

大规模数据的产生，对数据管理和处理技术提出了严峻的挑战。如何从各种各样类型的大量非结构化视频数据中精准提取目标信息，包括人、车、物以及各种行为信息，仍是视频信息处理的难点。未来的5G通信技术能够提供极快的传输速度，支持更多的终端接入，缩短延时时间。通信时间大幅缩短之后，“请求-响应”时间的长短对实时计算的要求更高，作为支撑智慧城市深入发展和广泛应用的核心技术，高性能计算也面临着机遇和挑战。

当前主流型号服务器，支持10—100通道视频实时处理，大城市监控视频达到10—100万个摄像头，后台处理就可能需要1万节点规模。

1.2 高性能计算在生命信息领域应用不断拓展

生物计算，包括DNA数据处理、蛋白质结构预测、脑模拟等，近20年来，呈现快速发展态势。这些计算大部分也是数据密集型的。

*（1）在基因数据处理应用方面。*高通量测序技术的进步，推动了基因数据爆炸性增长。然而，DNA包含信息的复杂性，仍对数据分析算法提出新挑战。机器学习在基因组分析与预测的多个方面获得研究与应用，如识别转录起始位点/剪切位点/TF结合位点、预测基因功能/疾病表型等[2]。深度学习应用的最新突破已经在预测性能方面超过了许多传统的统计推断算法，并且机器学习在基因组学中的癌症诊断、临床遗传、作物改良、流行病、公共卫生、人口遗传、进化、功能基因组等均具有明显发展潜力。当前，大多数问题的预测能力都没有达到实际应用的预期，对这些抽象模型的解释也没有阐明深刻的认识。如何有效利用机器学习获取更强大、更智能的基因组解释能力，还需要探索依赖于特定任务的机器学习模型。

*（2）在蛋白质结构预测应用方面。*蛋白质结构预测与设计对深入理解蛋白质结构和功能具有重要意义。蛋白质是一切生命系统的物质基础，但其生物功能的发挥，需要蛋白质正确折叠为特定的3D结构，蛋白质折叠研究也是药物设计的基础。实验学的方法如X射线（X-ray）、核磁共振（NMR）和冷冻电镜解析蛋白质3D结构普遍存在设备昂贵、时间和人力成本过高等问题。开发能够自动、快速、准确地将未知蛋白序列分类为特定折叠类别的计算预测方法成为计算生物学家长期努力的方向。使用深度卷积神经网络和残存网络高精度的预测蛋白中的氨基酸-氨基酸接触作用，并将预测结果直接用于蛋白质3D结构重建是近期的一个热点[3]。谷歌DeepMind将AlphaGo转型，开发了可预测蛋白质折叠的程序AlphaFold，并以该项目参加了全球蛋白质结构预测竞赛CASP13，取得了优异的成绩。麻省理工学院（MIT）的Belpler和Berger利用人工智能（AI）技术，直接通过氨基酸序列预测蛋白质分子的生物学功能。

*（3）在脑模拟和脑科学应用方面。*脑病变给人类带来的经济和生活负担远远超过其他领域，已成为全球致残的首要因素。2013年欧盟推出了15个欧洲国家参与、预期10年的“人类脑计划”，该计划侧重于借用超级计算机技术，通过研究脑连接图谱模拟脑神经网络功能。自该计划以后，其他国家纷纷提出各自脑计划。人脑大约由10^{10}个神经元和10^{14}个突触组成。现有的模拟研究多用于小区块模拟，规模约10^5个神经元。小规模模拟存在明显局限性，其中神经元连接的约50%突触分布在区块以外，而功能回路在整个大

ChinaXiv合作期刊

chinaXiv:202303.10253v1

脑内完成，区块之间相互影响明显。全脑模拟计算规模巨大，现阶段难以实现。当前，已达到的最大规模脑网络模拟是基于开源软件 NEST 完成的[4]。该网络包括 1.51×10^9 个神经元和 1.68×10^{12} 个突触，在日本超级计算机 K（共 88 128 节点，每节点 8 核，2 GHz，16 GB RAM）上全节点运行模拟。该模拟达到了人脑规模的 10%。

（4）*在新药创制应用方面*。高性能计算可在加速药物研发，降低药物开发风险的多个方面发挥重要作用，如超大规模筛选药物先导化合物、大规模搜寻药物潜在靶点、精确计算蛋白配体自由结合能、精确计算电子级别靶点药物共价和非共价作用、复杂生物体系模拟、药物网络化相互作用等。国内，北京大学、中国科学院上海药物研究所、上海交通大学在相关领域做出了可喜成果。

1.3 高性能计算在网络信息安全方面获得深入应用

网络信息安全已经成为国家安全的基石。近年来高性能计算技术也被广泛用于解决信息安全中面临的问题。典型应用场景，如模拟攻防对抗网络靶场、大数据隐私保护和入侵检测等。

（1）*网络靶场*。这是一种为网络技术、信息安全攻防技术和信息安全构想等提供定量和定性评估的实验环境，具备可信性、可控性和可操作性强等特点。目前，网络靶场已经在美国、英国、澳大利亚和日本等国实施。在网络靶场中，往往需要支持不同安全等级环境下的计算机网络防御、侦查、攻击测试，需要靶场具有对虚拟节点的可扩展能力，以生成成千上万的测试节点。例如，互联网环境生成技术 LARIAT[5] 可用于模拟互联网上的 HTTP、HTTPS、TCP/IP、SSH 和 SMTP 等协议；为了给网络靶场中成千上万节点生成的大规模网络流量，LARIAT 分布式地配置在每一个节点上。

（2）*隐私保护*。隐私保护是大数据时代面临的重大挑战。大数据技术往往具有两面性。一方面，研究人员可以通过大数据发现数据之间的内在关联，提供对事物的多视角洞察，给数据持有者更多的决策信息；另一方面，攻击者对大数据的内在联系的分析，往往使得数据的隐私难以得到有效的保障。在除去患者标记的某种疾病病例中，攻击者可以通过社保、出行记录、消费记录等数据来综合推断出患者的信息。因此，大数据中的数据隐私保护是至关重要的。

（3）*入侵检测*。入侵检测和防御系统有一定的实时性，因此后台对于数据处理的性能要求较高。Erfani 等[6]提出了一种深度置信网络和单边支持向量机混合的模型用于异常检测。从实验结果来看，采用高性能算法以后训练的时间减少为原来的 1/3，测试的时间减少为原来的 1/1 000。

1.4 传统高性能计算应用的数据驱动新方法获得成功实践

（1）*石油勘探领域新方法*。新兴的海量地震数据采集对地震数据处理、解释、建模等工作提出了极高的挑战，大规模数据处理理论与技术研究成为当前的研究热点。人工智能在石油勘探的应用研究主要分为3 类：解释、预处理、反演。① 地震数据解释是较早发展 AI 方法的领域，以 AI 代替人机交互为目的，解决大规模数据解释。传统地震解释属于人为工作，需要大量工作人员与专家参与。② 数据预处理以提高地震信号质量为目的，如数据重建、噪音压制、滤波等。传统预处理多利用积分变换方法，多存在常见的不理想现象。新兴的 AI 方法能够取得更好的预处理效果。③ 全波型反演是近期的研究热点，其迭代收敛导致计算量大幅度上升，且信号噪音的存在导致出现局部极小值收敛的问题。基于 AI 方法的反演成像，取得了一定进展。然而，AI 技术的应用对数据存储和计算资源也提出更高要求。

（2）*气象预报领域新方法*。数值预报是高性能计算的传统应用领域，几乎在任一时期，数值天气预报系统都使用了当时最快速的高性能计算机。AI 在

天气预报领域的应用有较长历史。当前，AI 技术在天气和气候领域中的应用研究主要包括：观测数据质量控制、卫星数据反演及数值模式资料同化、数值模式参数化及后处理、天气系统识别、极端和灾害性天气预报、短时临近预报、台风海洋天气预报、气候分析和预测、环境污染相关预报、可再生能源相关的预报等[7]。2013 年，Earth Risk 发布的 TempRisk Apollo 40 天的气温概率预报模式，利用深度学习方法填补了传统的数值天气预报 15—30 天之间的延伸期预报空白。2016 年，IBM 通过收购 The Weather Company（天气公司），实现 0.2—1.2 英里小尺度超局地天气的准确预报，精准服务于运输公司、公共事业单位甚至是零售商。2017年，中国气象局联合天津大学共同研发的全国强对流服务产品加工系统，运用图像识别和深度学习等新技术，判断出未来 30 分钟内强对流天气发生和影响的区域，并发布空间分辨率为 1 公里，每 6 分钟滚动更新的预测产品。然而，仅基于图像识别和深度学习的预报方法无法涵盖天气过程发生的物理基础，仍然无法取代传统数值模式预报。未来，随着计算能力提升和AI方法的进步，气象预报领域应用是否能有新的突破，值得气象学者与计算机学者共同协作，进一步探索。

2 高性能工业计算推动制造向智造迈进

计算机仿真用于模拟现实，并且用于探索新的理论，设计新的实验，以及测试新的理论。当现象无法观测，测量不切实际或者过于昂贵的时候，仿真为实验和观测科学提供了另一种选择。30 年来，仿真理论和技术对工程领域才产生了巨大的影响，计算仿真替代了大部分的物理实验，而且有一些无法做实验的极端工况可用计算机来仿真。计算机仿真的水平是衡量工业企业竞争力的主要标志之一，计算机仿真是成就未来工程和科学发展的关键。

国际上著名的制造业大公司已实现了产品的虚拟化设计和制造，并实现了全球资源共享。美国波音公司从 20 世纪 80 年代起，在飞机设计中利用计算机仿真替代了大量的风洞实验等传统的物理实验和样机验证过程。现在物理实验反而成为验证和辅助手段（图 1），飞机的设计周期大大缩短，研发费用大幅度下降。从飞机布局研究、关键气动部件设计、发动机设计到飞机性能分析，都广泛应用到了计算流体力学（CFD）技术。飞机全机计算中，采用工程湍流模式的全机网格规模早就突破千万量级。

汽车工业中，以计算机仿真为核心的数字化开发是重要的技术突破，也被认为是继福特流水线生产、丰田精益生产之后，汽车工业具有革命性意义的重大技术进步。以汽车安全性研究为例来看，早期的汽车耐撞性研究主要采用实验手段来进行，车对障碍物的撞击试验是评价一辆汽车抗撞强度的唯一方法。制造商为了评价某种车型零部件更改的效果并证明最终产品符合各项汽车安全标准，需要投入数百万美元的费用进行数百次试验，这是一种很昂贵的试错过程。从零部件直至整车装配级别的研发设计阶段都有大量计算分析，涉及刚度、强度、噪声、振动与声振粗糙度（NVH）、机构运动、碰撞模拟、板件冲压、疲劳和空气动力学分析等方面，计算机仿真发挥着无可替代的优势和作用，从而大幅度提高设计质量，缩短产品开发周期，节省大量开发费用。

上海汽车公司在十多年来“荣威”系列车型开发过程中，均借助了超级计算平台完成了大量虚拟安全碰撞试验计算工作，使虚拟碰撞试验数量、分析精度、精细程度和设计周期等都接近全球一流汽车研发水平。

3 新型应用发展趋势

新型行业应用的计算行为基本特征就是数据密集。大数据涌现在高性能计算应用广度和深度的进一步拓展方面、高性能计算系统发展方面、高性能算法

chinaXiv:202303.10253v1

ChinaXiv合作期刊

chinaXiv:202303.10253v1

和软件研发方面，带来一些新的趋势。

3.1 高性能计算应用广度和深度进一步拓展

（1）大数据应用层出不穷。在物联网技术、5G技术及各种数据采集技术快速发展的背景下，以大数据为基本特征的各类应用层出不穷，从智慧城市、生命健康到粒子物理、天体物理，大数据应用广度和深度不断拓展。

（2）传统问题的人工智能求解带来新机遇。人工智能产生应用的基本条件，大数据和计算力已经发展到新阶段，人工智能对传统问题，生命健康、基础物理乃至气象气候预报、石油勘探等众多领域，产生积极推动作用。2018年Gordon Bell奖之一，就是利用深度学习方法，分析气候变化模式[8]。

（3）多领域问题协同分析陆续出现。大数据的积累和计算力的大幅度提升，也为复杂系统或多领域问题协同分析创造了可能，如环境与生态的模拟问题、物理化学和生命的多尺度分析、精准医疗等。

3.2 高性能计算系统研发面临新的发展机遇

（1）专用加速硬件兴起。从传统的图形处理器（GPU）到张量处理器（TPU）、深度计算器（DCU）等专用硬件的流行，高性能计算加速芯片不断涌现。其中TPU[9]是近年来Google提出的专用于深度学习加速的张量处理单元，其提供低精度高通量的计算。相比GPU，TPU拥有更高的I/O（输入/输出）效率和更低的能耗。

（2）大数据处理对超级计算机网络提出新要求。典型的大数据处理问题，数据I/O等从存储到计算的频繁程度和带宽需求大幅度上升，而计算进程之间通信更多以小消息为主，约占消息总数的95%[10]。因此，大数据处理对存储网络和计算网络的性能需求，产生显著分离。

（3）大数据处理平台的非功能需求特点。新的业务对应用提出了众多非功能性的需求，如要求应用具备快速开发、可扩展、易重用、有统一接口、有完整的生态供下游业务进行作业，以及自动容错等。

3.3 高性能算法与软件设计面临新的挑战

（1）大数据矩阵的稀疏性特点。大多数大数据处理问题，最终转化成矩阵计算，这些矩阵通常

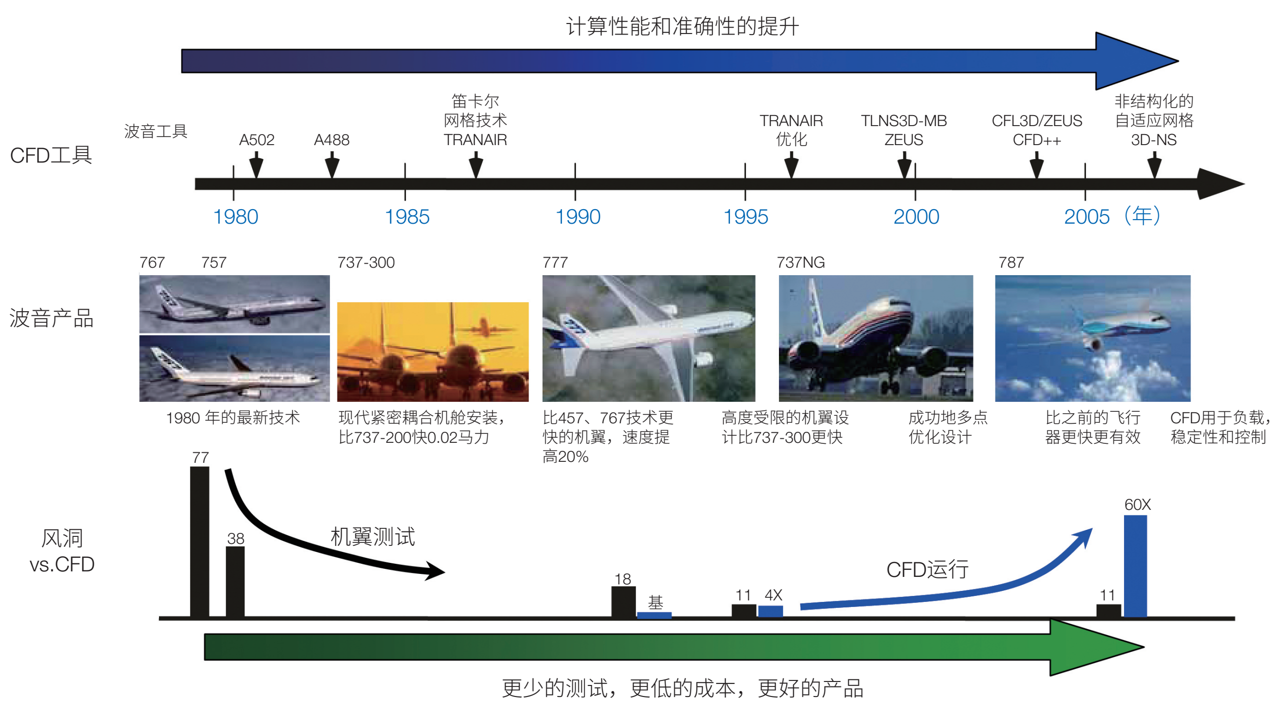


图1 新兴高性能计算行业应用及发展战略

chinaXiv:202303.10253v1

具有稀疏性（零元数据个数远远大于非零元数据个数），但一般不具备偏微分方程求解问题中稀疏矩阵非零元对角分布特性。另外，传统的稀疏矩阵求解通常要求精确解，但随着新应用的出现，人们对于稀疏矩阵的求解精度要求，往往让位于求解速度的要求。例如，在商品推荐、搜索引擎和社交网络等大规模推理任务中，用户往往只需要在 *Top k*（*Rank k*）的结果中出现想要的结果，即认为求解符合预期。而在求解性能方面，可能要求计算复杂度 $O(k)$（$k<<n$）的算法。

（2）*大数据时空性特点*。新兴智慧城市等应用中，时空特性越来越受到人们的关注。例如：交通流量预测中，GPS 前后有很强的时序关系；自然语言处理（NLP）中，上下文前后词语之间也呈现出明显的时空特性。

（3）*大数据高维度特点*。数据的维度通常随着业务问题规模的增大而指数级增大。为了约减数据集的维度，研究人员通常采用特征抽取、去不相关、去低方差和去常量属性等技术。数据体量大增加了数据维度约减的难度。

（4）*异构大数据融合特点*。大数据往往是异构的，数据的属性是不一致的。面临体量大、种类多和数据持续产生等特点。如何有效地融合多个异质数据集并挖掘出其中有用的信息，是高性能计算需要解决的问题。

（5）*大规模图处理特点*。基于图来发现事物之间的关联性是大数据领域的典型问题。大规模图处理往往面临数据局部性差、数据剖分困难、通信开销大等挑战，而实际应用对于算法实时性又往往存在较高要求。随着数据体量的增大，如何进一步减少数据的处理时间，是高性能计算面临的新挑战。

4 新型应用发展战略建议

针对上述新趋势，高性能计算发展面临新需求、新机遇，我们建议大力推动高性能计算系统、高性能计算环境、高性能计算应用等各个层面的创新，促进高性能计算在科技创新和国民经济社会发展的各个层面，发挥更大的作用。

4.1 加大高性能计算系统核心技术创新力度

（1）*加大专用加速硬件创新支持力度*。新兴应用对硬件加速计算提出了更高的 I/O 需求和更低的能耗要求，传统的众核加速处理器难以适应新兴应用需求，有必要加大专用众核加速处理器的创新与发展，为新兴应用提供更强大的高性能计算环境。

（2）*加大超级计算机网络创新力度*。为解决大数据存储与计算 I/O 次数和带宽需求大幅提升问题，有必要加大存储和计算网络的 I/O 性能的创新与发展，适应现有大数据处理对通信次数和粒度的需求。

4.2 构建面向新型应用的高性能计算环境

（1）*加大大数据系统平台创新力度*。为适应新兴应用对非功能性需求的提出，有必要发展大数据系统平台在可应用性、可扩展性、容错性的创新，建立更加完整的超算生态环境。

（2）*加大人工智能开放平台创新力度*。针对人工智能应用全流程，包括数据预处理、训练、参数调整、过程监控、测试等步骤，进行统一规划管理，研发构建领域数据集和智能开发一体化环境，有效降低非专业领域用户、科研用户使用难度，全面服务智能应用创新。

4.3 大力推进高性能应用软件研发

（1）*大力推进新型应用软件研发*。大数据、人工智能方兴未艾，智能技术被视为第四次技术革命的标志性技术，我们应大力推进智能制造、精准医疗、生命健康、智慧助理、智慧能源、智能驾驶、虚拟靶场等各类应用研发，促进新型应用更广更深入拓展。

（2）*大力推进高性能应用软件和工具软件研发*。目前高性能计算应用软件和工具软件占统治地位的是美国发展的，在该领域我国自主可控能力十分薄弱，

ChinaXiv合作期刊

chinaXiv:202303.10253v1

差距巨大，甚至远远落后于硬件的发展，风险极高，计算规模和复杂度也不够。需要持续推进高性能计算软件和工具软件研发。

4.4 大力推进传统应用的新方法开发

传统的计算化学、计算物理学、计算生物学等交叉学科仍在快速发展，药物智能筛选、分子动力学模拟等面临新的发展机遇，气候模拟、气象预报、能源开发、脑模拟、社会模拟等面临新的突破，大力推进数据驱动新方法研发具有重要意义。

4.5 大力推进大数据人工智能等新领域基准评测工具研发

新型应用数据呈现混合精度、数据密集与I/O密集叠加、通信行为的随机性增强，传统的以Linpack为代表的基准测试已经不能反映新型应用的计算需求，迫切需要研发能够代表新型应用计算行为特征的新领域基准评测工具。

参考文献

1 Reinsel D, 武连峰, Gantz J F, et al. IDC: 2025年中国将拥有全球最大的数据圈. Framingham: IDC, 2019.

2 Libbrecht M W, Noble W S. Machine learning applications in genetics and genomics. Nature Reviews Genetics, 2015, 16(6): 321-332.

3 Wang J X, Cao H L, Zhang J Z, et al. Computational protein design with deep learning neural networks. Scientific Reports, 2018, 8(1): 6349.

4 Jordan J, Ippen T, Helias M, et al. Extremely scalable spiking neuronal network simulation code: from laptops to exascale computers. Frontiers in Neuroinformatics, 2018, 12(2): 1-21.

5 Damodaran S K, Couretas J M. Cyber modeling & simulation for cyber-range events// Proceedings of the Conference on Summer Computer Simulation. San Diego: Society for Computer Simulation International, 2015: 1-8.

6 Erfani S M, Rajasegarar S, Karunasekera S, et al. High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning. Pattern Recognition, 2016, 58: 121-134.

7 McGovern A, Elmore K L, Gagne D J, et al. Using artificial intelligence to improve real-time decision-making for high-impact weather. Bulletin of the American Meteorological Society, 2017, 98(10): 2073-2090.

8 Kurth T, Treichler S, Romero J, et al. Exascale deep learning for climate analytics// Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis (SC). Dallas: IEEE Press, 2018: 649-660.

9 Jouppi N P, Young C, Patil N, et al. In-datacenter performance analysis of a tensor processing unit// ISCA’17 Proceedings of the 44th Annual International Symposium on Computer Architecture. Toronto: IEEE Press, 2017: 1-12.

10 Vanderbauwhede W, Benkrid K. High-performance Computing Using FPGAs. New York: Springer, 2013.

chinaXiv:202303.10253v1

Development Strategy of Emerging Applications of HPC

FENG Shengzhong[1*] LI Genguo[2] LI Xuelei[1] QI Fumin[1] HUANG Dian[1] WAN Yi[1] WU Jincheng[1]
（1 National Supercomputing Center in Shenzhen, Shenzhen 518055, China;
2 Shanghai Supercomputer Center, Shanghai 201203, China）

Abstract The rapid development of data-intensive emerging application areas in recent years is one of major characteristics of the extensive and in-depth applications of high performance computing (HPC). The HPC applications in the emerging areas bring new challenges and opportunities at all levels of HPC, including system technological innovation, computing environment innovation, and application innovation. Based on reviewing the application progress of HPC in the emerging areas, this paper summarizes the current technical characteristics and challenges, and provides strategic recommendations for the development of emerging HPC application areas, including increasing the core technology innovation of HPC systems, building HPC environment for emerging application areas, promoting the development of HPC application software and new methods for traditional applications, as well as promoting the development of benchmarking tools in new areas such as big data and artificial intelligence.

Keywords emerging application areas, data-driven, data-intensive computing, high performance computing (HPC)



冯圣中 国家超级计算深圳中心主任，中国科学院“百人计划”研究员、博士生导师。中国计算机学会杰出会员，科技部高性能计算重点专项总体组专家。曾作为骨干成员荣获中国科学院杰出成就奖、国家科技进步奖二等奖。主要研究方向为高性能计算、生物信息学等。E-mail: fengsz@nsccsz.cn

FENG Shengzhong Director of National Supercomputing Center in Shenzhen, Ph.D. Supervisor and Hundred Talent Program Professor of Chinese Academy of Sciences (CAS). He is a distinguished member of the China Computer Federation (CCF) and key special group expert of high-performance computing of the Chinese Ministry of Science and Technology. He won the Outstanding Achievement Award of CAS and the Second Prize of National Science and Technology Progress Award as a backbone member. His primary research interests include high-performance computing and bioinformatics. E-mail: fengsz@nsccsz.cn

■责任编辑：岳凌生

*Corresponding author

ChinaXiv合作期刊

chinaXiv:202303.10253v1

参考文献（双语版）

1 Reinsel D, 武连峰, Gantz J F, et al. IDC: 2025年中国将拥有全球最大的数据圈. Framingham: IDC, 2019.

Reinsel D, Wu L F, Gantz J F, et al. IDC: China will have the largest data circle in the world in 2025. Framingham: IDC, 2019. (in Chinese)

2 Libbrecht M W, Noble W S. Machine learning applications in genetics and genomics. Nature Reviews Genetics, 2015, 16(6): 321-332.

3 Wang J X, Cao H L, Zhang J Z H, et al. Computational protein design with deep learning neural networks. Scientific Reports, 2018, 8: 6349.

4 Jordan J, Ippen T, Helias M, et al. Extremely scalable spiking neuronal network simulation code: From laptops to exascale computers. Frontiers in Neuroinformatics, 2018, 12(2): 1-21.

5 Damodaran S K, Couretas J M. Cyber modeling & simulation for cyber-range events// Proceedings of the Conference on Summer Computer Simulation. San Diego: Society for Computer Simulation International, 2015: 1-8.

6 Erfani S M, Rajasegarar S, Karunasekera S, et al. High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning. Pattern Recognition, 2016, 58: 121-134.

7 McGovern A, Elmore K L, Gagne D J, et al. Using artificial intelligence to improve real-time decision-making for high-impact weather. Bulletin of the American Meteorological Society, 2017, 98(10): 2073-2090.

8 Kurth T, Treichler S, Romero J, et al. Exascale deep learning for climate analytics SC18: International Conference for High Performance Computing, Networking, Storage and Analysis. Dallas, TX, USA. IEEE, 2018: 649-660.

9 Jouppi N P, Young C, Patil N, et al. In-datacenter performance analysis of a tensor processing unit// ISCA’17 Proceedings of the 44th Annual International Symposium on Computer Architecture. Toronto: IEEE Press, 2017: 1-12.

10 Vanderbauwhede W, Benkrid K. High-Performance Computing Using FPGAs. New York: Springer, 2013.